

## 1 A3C の擬似コード

## 2 Atari Games の環境のラップ処理

Deep Q Network では Atari Games から受け取った状態の前処理などを行うことで学習をしやすくしている。以下にそれらの処理を示す。なお,  $reset_{\text{hoge}}()$  は環境をリセットするときに呼ぶメソッドを,  $step_{\text{hoge}}(a_t)$  は行動  $a_t$  を取って環境を更新するときに呼ぶメソッドを,  $observe_{\text{hoge}}()$  は環境の状態を返すときに呼ばれるメソッドを,  $reward_{\text{hoge}}()$  は報酬を返すときに呼ばれるメソッドを表す。

---

**Algorithm 1**  $reset_{\text{noop}}()$ 

---

**Input:**  $env, l_{\text{nomax}}$

- 1: エピソードの開始時に, 数フレーム何もしない行動を取り, 初期状態を決定する.
- 2:  $T \sim U(1, l_{\text{nomax}})$
- 3: **for**  $t' = 1, \dots, T$  **do**
- 4:    $a_{t'} = (\text{do nothing})$  の実行
- 5: **end for**

**Output:** 初期状態が決定した環境  $env$

---

---

**Algorithm 2**  $step_{repeat}(a_t)$ 

---

**Input:**  $env, l_{repeat}, a_t$ 

- 1: 1 回行動を取ると、同じ行動を指定フレーム続ける。指定数分行動を繰り返したら、直前のフレームの観測との最大値を状態として返す。
- 2: ※  $a_t$  は選択したい行動とする。
- 3:  $r_{total} = 0$
- 4: **for**  $t' = 1, \dots, l_{repeat}$  **do**
- 5:    $s_{prev} = s_{t'}$
- 6:    $a_{t'} = a_t$  として行動を選択し、環境  $env$  を更新、 $(s_{t'+1}, r_{t'}, done_{t'})$  を観測する。
- 7:    $r_{total} = r_{total} + r_{t'}$
- 8:    $s_{max} = \max(s_{prev}, s_{t'+1})$
- 9:    $done = done_{t'}$
- 10: **end for**

**Output:**  $s_{max}, r_{total}, done$ 

---

---

**Algorithm 3**  $observe_{gray84}()$ 

---

**Input:**  $s_t$ 

- 1: 観測した画面を (84,84) サイズのグレースケール画像に変換して返す。
- 2:  $s_t$  をグレースケール画像に変換
- 3: 変換後の  $s_t$  をさらに (84,84) に reshape

**Output:** 変換後の  $s_t$ 

---

---

**Algorithm 4**  $step_{stack}(a_t)$ 

---

**Input:**  $env, l_{history}, a_t, S$ 

- 1:  $l_{history}$  ステップ数分の観測の履歴を状態として返す。
- 2:  $S$  は観測の履歴とする。
- 3:  $a_t$  を行動として選択し、環境  $env$  を更新、 $(s_{t+1}, r_t, done_t)$  を観測する。
- 4:  $S$  に  $s_{t+1}$  を追加。  $|S| > l_{history}$  なら、一番古い履歴を  $S$  から削除する。

**Output:**  $S, r_t, done_t$ 

---

---

**Algorithm 5**  $reward_{clip}()$ 

---

**Input:**  $r_t$ 

- 1: 報酬  $r_t$  が正なら +1 に、負なら -1 に、0 なら 0 として返す。
- 2: **if**  $r_t > 0$  **then**
- 3:    $r_t = 1$
- 4: **else if**  $r_t < 0$  **then**
- 5:    $r_t = -1$
- 6: **else**
- 7:    $r_t = 0$
- 8: **end if**

**Output:**  $r_t$ 

---